Early Detection & Classification of Diabetic Nephropathy Using Machine Learning Techniques.

Miss. Sujata V.Kale[1]
Assistant Processor Dept. of Computer Science
DayanandScienceCollege,Latur,Maharashtra (India)
sujata1010@gmail.com

Mr.Mangesh B. Sugare[2]
Assistant Processor Dept. of Computer Science
DayanandScienceCollege, Latur,Maharashtra (India)
mangesh.sugare@gmail.com

Abstract: Diabetic nephropathy is a common disease of type-1 diabetes and type-2 diabetes. It is a usual problem and main cause of death in people with diabetes. Uncontrolled diabetes can damage the blood vessels in kidneys so that filter of the waste in your blood can not be done properly. This will lead to kidney damage and high blood pressure .The high blood pressure can cause further damage to the kidneys by increasing the pressure in the delicate filtering system of the kidneys.The complications of diabetic nephropathy may develop gradually over months or years. In this work study of ensemble algorithm included Bagging, AdaBoost and Random Forest, Gradient Boosting,Bayesian Networks technique is done

Keywords: Diabetes, Diabetic Neuropathy, Bagging, AdaBoost and Random Forest, Gradient Boosting,Bayesian Technique, Diabetic kidney disease (DKD).

1 Introduction

Diabetic kidney disease is the leading reason of death in diabetic patients and accordingly manifests itself in almost half of patients diagnosed with type 2 diabetes.Diabetic nephropathy kidney disease that results from diabetes is the number main cause of kidney failure. Diabetic Nephropathy is in near about a third of people diagnosed with diabetes .Diabetic nephropathy is a serious cause of chronic kidney disease and at endresults in renal failure globally. Diabetic nephropathy or diabetic kidney disease is a syndrome expressed by the presence of pathological quantities of urine albumin excretion, diabetic glomerular lesions, and loss of glomerular filtration rate (GFR) in diabetics [1].Most people in the early stages of DKD either show non-specific symptoms or have no symptoms, contributing to a misdiagnosis. At the time of diagnosis, patients do not have an accurate initial estimate of the risk of developing Diabetic kidney disease.Diabetic nephropathy is a major cause of long-term kidney disease. So the kidneys no longer work well enough to meet the needs of daily life. So it can lead to kidney failure with potentially life-threatening consequences.

2. Literature and Review

Paper	Algorithm	accuracy	Sensitivity	Specificity
1 Chronic Kidney Disease Prediction using Machine Learning Ensemble Algorithm[9]	Bagging	0.991666	1	0.988095
	AdaBoost	1	1	1
	Gradient Boosting	0.983333	0.9759	1
	Random Forest	1	1	1
2. Extraction of Action Rules for Chronic Kidney Disease using Naïve Bayes Classifier[10]	best first search	97.75%	-------	----------
	Naïve Bayes	80%	50%	99%
	Naïve Bayes ONE R	98%	95%	100%

2.1 Chronic Kidney Disease Prediction using Machine Learning Ensemble Algorithm: In this work Bootstrap integration is commonly known as bagging a kind of ensemble algorithm that selects something randomlyExamples from the training set with replacement [2]. In Bagging, bootstrap samples are taken from the training. The results of each classifier are aggregated and finalized. The result is a majority voting process. Tests show that bagging can be used to upgrade .The overall performance of a possibly weaker classifier [3].

AdaBoost works by selecting base classifiers andimprove its efficiency by identifying incorrectly classified casesRepetition methods from the training dataset. After each iteration, base classifiers .The trainings are applied to the dataset and increase its weight incorrectly classified features. The cycle is repeated n times, Applying a base classifier to the training set each time Updated weight.Random forest helps in clinical applications for betterAccuracy by combining a bunch of weak classifiers like decision tree. The decision by creates the N number of treesUsing randomly selected attributes as their information [3].

Gradient Boosting model is Creates decision tree using reasonable gradients to minimize Damage function. The final estimate is made using weights. The dominant part of the whole decision tree opinion [4] [5] [9].

2.2 Extraction of Action Rules for Chronic Kidney Disease using Naïve Bayes Classifier However, it is known that if a data set with 30 variables has a 5% value reduction (randomly spread across all points and records), then one would have to omit the data set approximately 80% of the record. Missing numerical properties are the mean or mid-range attributes, as its name implies, replacing the lost values in the record with the median value of that attribute taken in the entire data set. However for nominal attributes Mode Imputation is done, replaced the missing values in the record with the mode value of that attribute taken across the data set [6].

Feature selection is a widely used to reduce thedimensionality by extracting small subsets of relevant featuresfrom the original data based on high relevance assessment

standard. OneR is a short, simple but accurate, classification algorithm for "one rule" that generates a rule for each predictor in the data and selects the rule with the smallest total error as "one rule". [7] [8].Chronic Kidney Disease is predicted by using Naïve Bayesclassifier. Naive Bayes classifiers are greatly extensible, which require a number of parameters linear in the number of variables (features/predictors) in a learning problem. Different stages are predicted by calculating GFR from thegiven attributes using MDRD formula. GFR is traditionally measured as renal clearance of an ideal filtration marker, suchas inulin from plasma. Estimating GFRbased on a filtration marker (usually serum creatinine) is nowwidely accepted as an initial test. By considering the predicted stage from the former stepaction rules are generated with respect to the stage [10].

3. Conclusion:

Diabetic Kidney Disease (DKD) is the diseases that affect the people in large numbers. As the symptoms of DKDare not visible in the early stages many a times the disease isonly detected when it has reached an advanced stage.The four-ensemble algorithm like Bagging, Random Forest, AdaBoost and Gradient Boosting were used. Based on Accuracy the AdaBoost and Random Forest performed better with 100% Accuracy. Based on Precision, Bagging showed 97.29% and AdaBoost, Gradient Boost and Random Forest showed 100%. According to evaluation AdaBoost and Random Forest was the best classifier among the Bagging and Gradient Boosting. This the risk of nephropathy is minimized in by comparing the effect of different risk factors. It presented the power for diagnosis and prevention diabetes complications. This will enable it to apply this technique in the decision support tool for avoidance and betterment of the quality of life for diabetic patients.

We have proposed Nave Beige with OneR number Attributes in datasets using OneR have also been reduced by 80% Algorithm and comparison accuracy improved by 12.5% to the present system. Our proposed system works So that the rules of the relevant chronic renal disease stage Necessary treatment can be done as per the rules of action Asked CKD to avoid going to the next level

4. References

1. Lim A. Diabetic Nephropathy–Complications and Treatment. *Int J Nephrol Renovasc Dis*.2014;7:361-381, https://doi.org/10.2147/IJNRD.S40172
2. C. B. C. Latha and S. C. Jeeva, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,” Inform. Med. Unlocked, vol. 16, p. 100203, Jan. 2019, doi: 10.1016/j.imu.2019.100203
3. I. Yekkala, S. Dixit, and M. A. Jabbar, “Prediction of heart disease using ensemble learning and Particle Swarm Optimization,” in 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Aug. 2017, pp. 691–698, doi: 10.1109/SmartTechCon.2017.8358460.
4. A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, “A datadriven approach to predicting diabetes and cardiovascular disease with machine learning,” BMC Med. Inform. Decis. Mak., vol. 19, no. 1, p. 211, Nov. 2019, doi: 10.1186/s12911-019-0918-5
5. R. Islam and Md. A. Shahjalal, “Soft Voting-Based Ensemble Approach to Predict Early-Stage DRC Violations,” in 2019 IEEE 62nd International Midwest

Symposium on Circuits and Systems (MWSCAS), Aug. 2019, pp. 1081–1084, doi: 10.1109/MWSCAS.2019.8884896.
6. Jiliang Tang, Salem Alelyani and Huan Liu, “Feature Selection for Classification: A Review”
7. Yvan Saeys, Inaki Inza and Pedro Larranaga, “A review of feature selection techniques in bioinformatics”, Bioinformatics, Volume 23, Issue 19, August 2007, pp 2507-2517.
8. Zheng, Lijuan, Hongwei Wang, and Song Gao, "Sentimental feature selection for sentiment analysis of Chinese online reviews", International Journal of Machine Learning and Cybernetics, 2015
9. Nikhila, "Chronic Kidney Disease Prediction using Machine Learning Ensemble Algorithm," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 476-480, doi: 10.1109/ICCCIS51004.2021.9397144
10. U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve bayes classifier," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016, pp. 1-5, doi: 10.1109/ICCIC.2016.7919649.
11.